REMARKS

The above amendment with the following remarks is submitted to be fully responsive to the Office Action of October 5, 2004. Reconsideration of this application in light of the amendment, and the allowance of the pending claims are respectfully requested.

Claims 1-21 and 23-34 were pending in the present application prior to the above amendment. In response to the Office Action, claims 1, 2, 4, 5, 7, 9, 17, 18, 20, 21 and 32 have been amended above, and claims 3 and 8 have been canceled. Therefore, claims 1, 2, 4-7, 9-21, and 23-34 are still pending in the present application and are believed to be in proper condition for allowance.

The present invention is directed to a method and system for extracting data of interest from a plurality of web sites. Presently, crawlers are created by computer programmers to retrieve information from a particular web site, for example, to extract desired information for a category of products from on-line merchants for use in an electronic catalog. However, it should be understood and appreciated by the Examiner that different web sites, for example, web sites of different on-line merchants, utilize different data structures. There is no standardized structure, method or protocol for presenting and storing information or data among different web sites that is uniformly followed by different on-line merchants. In addition, each web site generally utilizes a plurality of web pages in the web site to which a user has to navigate to obtain the desired data of interest regarding a product available through the web site, for example.

A crawler that is created and used to extract data from one web site generally cannot be used to extract data from other web sites due to the variations in data structure, method and/or protocol implemented by other web sites. Thus, a new crawler must be created by a computer programmer to extract data for each web site, the creation of new crawlers being time consuming and expensive. Consequently, extracting data of interest, for example, regarding a particular product from a plurality of different web sites such as merchant web sites, can be extremely difficult, expensive, and time consuming.

The present invention provides a novel method and system for extracting data of interest from a plurality of web sites that greatly facilitates the extraction process by providing tools that can be used, even by non-programmers, to extract desired information from the plurality of web sites. More specifically, the present invention allows the user to generate extraction patterns directly from the output from the web site itself so that other desired information can also be extracted from the web site. In accordance with the preferred embodiment, a respective description of data of interest is created for each web site that identifies the web site, an extraction pattern that extracts information from the respective web site is developed based on output from the respective web site, and the developed extraction pattern is associated with the respective description of data of interest for the web site. In addition, a value that can be used as an extraction parameter for the developed extraction patterns is received, and the desired data of interest is obtained by querying the plurality of web sites using the value and the extraction patterns associated with the respective descriptions of data of interest. If the data of interest includes data of interest from at least two web sites, the data of interest from the two web sites is provided.

Thus, as described in the Specification of the application, the present invention can be used by an individual such as a programmer, or even a non-programmer, to generate extraction patterns easily based on the output of the web site itself. A value can then be used in conjunction with the extraction pattern to extract different data of interest from the particular web site. Correspondingly, the present invention allows facilitated extraction of desired information from a plurality of web sites in a rapid, cost effective manner, without requiring a programmer to create a crawler for each web site. For example, the present invention can be readily used to improve comparative shopping experience for consumers by providing tools that a shopping agent can use to provide up-to-date information from a plurality of on-line merchants.

Referring now to the Office Action, the Examiner rejected claims 1-19 and 21-34 under 35 U.S.C. 103(a) as being unpatentable over U.S. Patent No. 4,992,940 to Dworkin. The Applicants respectfully disagree for the reasons set forth herein below

and request reconsideration and withdrawal of this rejection with respect to the now pending claims 1, 2, 4-7, 9-21, and 23-34.

Dworkin reference relates to a system and method for automated selection of equipment for purchase where the user selects a category of product or service, and the user is provided with a template which gives various criteria for the product or service selected. Dworkin discloses that upon receiving input from the user as to one or more criteria, the system searches the database for all products that fulfills the requirements inputted by the user. Dworkin also discloses that the database includes information regarding products from a plurality of vendors or distributors within the selected category. The results of the search are displayed for the user identifying the products together with the vendor. Thus, the cited Dworkin reference is essentially a multi-vendor catalog or search engine that provides the user with information regarding the product desired, in conjunction with the identity of the vendor.

The invention described in Dworkin is substantially different from the present invention. In this regard, it is important to note that Dworkin assumes that the database is populated with information regarding products that are available from plurality of vendors. Dworkin is silent as to how this database is populated or as to how the searching of the databases of different vendors are attained. Thus, Dworkin presumes that the databases of different vendors can be searched to extract the information regarding the products available, and to store such extracted information in the database that is searched based on user inputted criteria to find the desired matches.

However, as noted above, different web sites utilize different data structures, and each web site typically utilize a plurality of web pages in the web site which would require the user to navigate through various web pages to obtain/extract the desired data. Dworkin is silent as to how this database of information regarding products from a plurality of vendors can be attained. Correspondingly, without explicit teaching to the contrary, one of ordinary skill in the art would understand that the conventional method of using crawlers would be implemented in the system and method disclosed in Dworkin. Consequently, Dworkin does not contribute to solving the problem of

requiring individual, web site specific crawlers to be created by a programmer to extract the product information required to populate the database used by the system and method of Dworkin. Again, the creation of new crawlers for each web site would be time consuming and expensive.

In contrast, the present invention provides a method and system for extracting the desired information from a plurality of web sites so that, for example, database such as that noted in Dworkin, can be populated with information in a cost effective, efficient manner, without requiring creation of crawlers for each different web site. In contrast to Dworkin which is a front end system and method for providing a consumer with information regarding a particular product or service from a database, the present invention is a back end system and method for extracting desired data from a plurality of web sites for use in, for example, shopping assistants.

As explained in response to the previous Office Action, Dworkin does not disclose, teach, or otherwise suggest development of an extraction pattern recited in the rejected claims of the present application. Dworkin does disclose a template and a user inputting criteria into the template, the system searching the database based in the inputted criteria. In this regard, it may be argued that the inputted criteria disclosed in Dworkin is analogous to the received value recited in the present independent claims 1, 18, 21, and 32. However, Dworkin does not disclose an extraction pattern based on output of a web site, the extraction pattern being adapted to extract information from the respective web site.

In the above regard, it cannot be reasonably argued that the template of Dworkin is equivalent to the extraction pattern recited in the present claim because the template is predefined and provided to the user, and is not developed based on the output from the respective web site. In this regard, Dworkin is silent as to how the template is initially derived. Moreover, as noted above, Dworkin presumes that the template will work to obtain the desired information from the database, again presuming an existing database that is already populated with product information. Dworkin discloses that the database is queried using the template and the inputted criteria. Correspondingly, without specific teaching to the contrary, it would be

evident to one of ordinary skill in the art that Dworkin discloses conventional searching techniques using declarative queries of a structured query language to search an existing populated database.

In contrast, in the present invention, the extraction patterns are developed based on the output of the web site itself. Thus, the extraction patterns are not predefined, at least until they are developed using the output of the web site. In addition, a plurality of web sites are queried using the value and the developed extraction patterns to extract the desired information. Conventional structured query language cannot be readily used to extract desired information from web sites because, as previously noted, most web sites include a plurality of web pages that need to be navigated to obtain the desired information. Finally, Dworkin also does not disclose that when the data of interest includes data from two or more web sites, the data from the web sites are provided. This is because, as noted, Dworkin works within the database provided.

In response to the previously filed amendment of June 21, 2004, the Examiner asserts that the present invention sets forth an HTML web product and service search engine tool using standard database software tools and programming software tools. This assertion is incorrect in that the extraction tool as described and claimed in the present application is uniquely customized for the purpose of extracting data of interest from a web site by developing an extraction pattern using the output of the web site itself. As described, the preferred embodiment of the present invention provides a graphical user interface tool to facilitate development of an extraction pattern. Once the extraction pattern has been developed for the web site, new values indicative of the desired data of interest are used in conjunction with the developed extraction pattern, to extract the desired information from the web site.

In addition to the above, the Applicants respectfully contend that the Examiner's reliance on Dworkin and summary assertions as to obviousness based on databases and programming languages is improper, and does not establish a *prima facie* case of obviousness. These assertions of the Examiner are made without properly establishing any basis for combining the features of Dworkin with other known search techniques for databases in the manner proposed by the Examiner. In

this regard, Examiner appears to be engaging in improper hindsight reconstruction based on the present invention to obtain the required motivation for combining the references or teachings to assert that the present invention is "obvious", without properly citing references or any teachings in the prior art as to the features of the present invention recited.

The impropriety of the Examiner's summary rejection is evidenced by the fact that the Examiner has taken twelve "official notices" in rejecting the pending claims as being obvious, when use of such notices should be rare and judiciously applied, especially when an application is under final rejection. (See MPEP 2144.03). For example, the Examiner asserts that Dworkin teaches providing a tool for creating a program to extract data using at least one extraction parameter. While the Examiner further admits that Dworkin does not teach the web site, he takes official notice that link construction is well known and that it would be obvious to implement this feature "for the advantage of increased revenue by greater exposure to on-line customers and products." This statement reveals that the Examiner is not fully appreciating the present invention which is directed to extraction of information from web sites, and not to a system used by on-line customers. In addition, as discussed above, Dworkin merely disclose a database search tool for searching a pre-existing database whereas the present claims recite extraction patterns developed using the output of the web site itself. This is not suggested by Dworkin, by web sites generally, or by link construction in HTML. The Examiner taking notice of existence of web sites does not address the deficiencies of the rejection. Correspondingly, even if Dworkin is modified with the teaching officially noted by the Examiner, the modified Dworkin reference still fails to result in the present invention.

In another example, the Examiner rejects claim 2 noting that while Dworkin does not disclose a graphical user interface tool with a web browser for developing the corresponding program, and taking another official notice that providing a graphical user interface tool including a web browser is known. Based on this official notice, the Examiner asserts that it would have been obvious to provide a graphic interface tool for developing the program. However, the basis for this official notice is not

understood and the Applicants specifically request a reference showing such a feature. Moreover, as explained, Dworkin does not disclose developing an extraction pattern based on output from the respective web site. Correspondingly, even if Dworkin is modified with the teaching officially noted by the Examiner, the modified Dworkin reference still fails to result in the present invention.

Despite the deficiencies of the Office Action noted above, to further expedite the prosecution of the present application, independent claim 1 has been amended to specifically recite that an extraction pattern is developed based on output from the respective web site using a graphical user interface tool, the extraction pattern identifying at least a portion of the output of a web site and extracting information from the respective web site. The provision of the graphical user interface tool further facilitates developing of the extraction pattern by displaying the output from the respective web site, and allowing the user (such as a non-programmer) to graphically select the portions of the displayed output that is to be used as the extraction pattern as explained above. This feature and functionality of the present invention as now claimed are shown, for example, in Figures 15 to 17 and in the corresponding discussions in the Specification of the application. This limitation was set forth in claim 2, and has been examined by the Examiner. Thus, the added limitation does not raise any new issues and the entry of the added limitation is respectfully requested. Clearly, the prior art of record as discussed above, fail to disclose, teach, or otherwise suggest the method as now claimed. Therefore, the reconsideration and allowance of claim 1, as well as dependent claims 2, 4-7, 9-17, 23-26, and 29 that are ultimately dependent on claim 1, are respectfully requested.
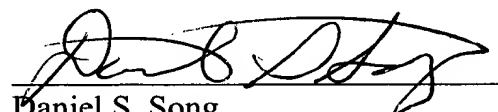
Similarly, independent claim 18 has been amended to specifically recite a means for developing an extraction pattern based on output from the web site, the means including a graphical user interface tool, the extraction pattern extracting data from the web site. Therefore, the reconsideration and allowance of claim 18, as well as dependent claims 19-20, 27, and 30 that are ultimately dependent on claim 18, are respectfully requested.

W342664.2

Likewise, independent claims 21 and 32 have also been amended to specifically recite a graphical user interface tool, and that the extraction pattern identifies at least a portion of the output of a web site, and extracts information from a web site. Therefore, the reconsideration and allowance of claims 21, 28, and 31-34 are also respectfully requested.

Finally, to the extent to which the Examiner disagrees with the remarks and reasoning set forth above, the Applicants respectfully request that the Examiner cite specific references, and teachings or suggestions for combing such references, to properly establish a *prima facie* case for obviousness instead of summarily asserting obviousness.

In view of the foregoing, it is submitted that the present application is in condition for allowance and a notice to that effect is respectfully requested. However, if the Examiner deems that any issue remains after considering this response, he is invited to call the undersigned to expedite the prosecution and work out any such issue by telephone.

Respectfully submitted,

Daniel S. Song
Registration No. 43,143

NIXON PEABODY LLP
401 9th Street, N.W., Suite 900
Washington, D.C. 20004-2128
(202) 585-8000
(202) 585-8080 (Fax)
Customer No. 22204

Date: December 2, 2004

W342664.2